

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 June 2001 (07.06.2001)

PCT

(10) International Publication Number  
WO 01/40301 A2

(51) International Patent Classification<sup>7</sup>: C07K 14/47

[GB/GB]; 7 Pitsligo Road, Edinburgh, Central Scotland  
EH9 1EG (GB).

(21) International Application Number: PCT/EP00/11915

(74) Agents: DE WEERD, P. et al.; P.O. Box 20, Wethouder  
van Eschstraat 1, NL-5340 BH Oss (NL).

(22) International Filing Date:  
28 November 2000 (28.11.2000)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
99309667.6 1 December 1999 (01.12.1999) EP

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (for all designated States except US): AKZO  
NOBEL N.V. [NL/NL]; Velperweg 76, NL-6824 BM  
Arnhem (NL). MEDICAL RESEARCH COUNCIL  
[GB/GB]; 20 Park Crescent, London W1N 4AL (GB).  
UNIVERSITY OF EDINBURGH [GB/GB]; Old Col-  
lege, South Bridge, Edinburgh, Central Scotland EH8 9YL  
(GB).

Published:

— Without international search report and to be republished  
upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 01/40301 A2

(54) Title: A NOVEL GENE, DISRUPTED IN SCHIZOPHRENIA

(57) Abstract: A newly identified gene, DIS1 is disrupted by a (1;11)(q42.1;q14.3) translocation which segregates with schizophrenia. We have examined the genomic structure of DIS1 and found that the gene consists of 13 exons estimated to extend across at least 300kb of DNA. Exon 11 contains an alternative splice site which removes 66 nucleotides from the open reading frame. The final intron of DIS1 belongs to the rare AT-AC class of introns. 8 expressed sequence tags (ESTs) located within introns 3, 7, 9 and 10 of DIS1 have also been identified. These ESTs have not yet been assigned to DIS1 and may therefore represent further novel genes in the region.

## A novel gene, disrupted in schizophrenia

The present invention relates to a newly identified DNA sequence which surrounds a breakpoint on chromosome 1 involved in a balanced t(1;11)(q42.1;q14.3) translocation as well as to a gene disrupted by this translocation event and its encoded proteins as well as to antibodies thereto and their use as a medicament. The invention also relates to methods for the detection of the translocation event as well as to methods for the in vitro diagnosis of a psychiatric disorder. Moreover, the invention also relates to transformed cell lines.

10

Family, twin and adoption studies have convincingly demonstrated a significant genetic contribution to schizophrenia (1995, Lancet 346: 678-682, and references therein) and have driven studies directed at identification of this genetic component. Schizophrenia is a complex disease and the multifactorial and probable genetic heterogeneity of the condition complicates the application and interpretation of conventional linkage and association studies. At present, however, no specific genes have been described which could play a role in schizophrenia.

Previously, a balanced t(1;11)(q42.1;q14.3) translocation was reported which is linked to schizophrenia and other related mental illness in a large Scottish family (1990, Lancet 336: 13-16) with a maximum LOD of 6.0 (Douglas Blackwood, in preparation). Mapping of the translocation breakpoint on chromosome 11, and the accompanying search for neighbouring genes has already been reported (1997, Am. J. Med. Genet. 74: 82-90, 1998, Psychiatr. Genet. 8: 175-181). No evidence for the presence of any part of a gene closer than 250kb to the breakpoint has been found.

It will be clear that there is a great need for the elucidation of genes related to schizophrenia in order to unravel the various roles these genes may play in the disease process. A better knowledge of the genes involved in schizophrenia and the mechanism of action of their encoded proteins might help to create a better insight into the etiology of this psychiatric disorder and its underlying molecular mechanisms. This could eventually lead to improved therapy and better diagnostic procedures.

The present invention provides such a novel gene which is located on chromosome 1 and is directly disrupted by the translocation event. More specific, the present invention provides for a gene, tentatively called DIS1 (Disrupted In Schizophrenia) 5 whose cDNA sequence is shown in SEQ ID NO:1. DIS1 is disrupted within an intron of the gene with the result that a proportion of the coding sequence has been translocated to chromosome 11.

The protein encoded by DIS1 is predicted to have a globular N-terminal domain(s) 10 and a helical C-terminal domain with the potential to interact with other proteins via formation of a coiled coil. The coiled-coil structure is present in several proteins (particularly microtubule binding proteins) which are involved in the development and functioning of the nervous system. The putative structure of DIS1 is therefore compatible with a role in the nervous system.

15 DIS1 consists of 13 exons which we estimate to extend across at least 300kb of genomic DNA. The translocation breakpoint lies within intron 8 of this gene. The effect of the translocation is therefore to remove exons 9-13 to chromosome 11. There is a commonly used alternative splice site, which does not disrupt the open 20 reading frame, within exon 11 which give rise to two distinct polypeptides as provided in SEQ ID NO: 2 and SEQ ID NO: 3. Table 1 shows the nucleotide sequences of the splice sites. The sequence of intron 8 is now revealed I SEQ ID NO: 4. At nucleotide 8432 a gap of unknown size occurs in the sequence.

25 The density of genes in the chromosome 1 breakpoint region is apparently high since, in addition to DIS1, 8 independent ESTs have also been identified. While this may suggest the presence of other genes in the region, it is also possible that some of these ESTs represent differentially spliced exons of DIS1.

**Table 1** DIS1 splice site sequences

exon	exon size	position	splice acceptor	splice donor
1	120bp	1-120	N/A	CACCGCGCAGgttagggagc
2	980bp	121-1100	ttcttccagGCAGCCGGGA	GCAGATGGAGgtcagtgtct
3	70bp	1101-1170	accaacatagGTAATATCCT	TATGATAAAAGgtgagttta
4	151bp	1171-1321	gggttccagCTGAGACGTT	CCACTCAGCAgtgaatacct
5	130bp	1322-1451	ttgtttaagGCCAGCGGA	GCAGCTACAGgtgagcaggt
6	236bp	1452-1687	ttctctacagAAAGAAATTG	CCATAAGGAGgtactgctga 1
7	55bp	1688-1742	attttccagCCTCCAGGAA	CACTACTAAAGgtaaagtacct
8	103bp	1743-1845	ctccccctagGTGTGTATGA	GCCATATCAGgtaaactggca 1
9	189bp	1846-2034	cgtgtgttagGAAACCATT	ACTGCCTATGgttaggtatgt
10	61bp	2035-2095	ttttcccccagAAACAAGTGT	AACTGTGCAGgtaaaggataa 1
11a	199bp	2096-2294	tctgtctcagCTGCAAGTGT	CCCTTGAAAGgtattggaaag
11b	265bp	2096-2360	tctgtctcagCTGCAAGTGT	ACAGAAAGAGgtctgtccct
12	118bp	2361-2478	ctctcgccagGAATCTTACA	GATCTCATTCatatcccttt 1
13	<u>4430-4585*</u>	2479-6913	ctcctaacaatgtgcccacAGTCTCTCAG	N/A

Exon size depends upon poly(A) signal usage and poly(A) addition site selection

DIS1 is predicted to encode a protein with an N-terminal globular head consisting

5 mainly of beta-sheet, and solvent-exposed helical tail with the potential to form coiled-coils. The transition from beta-sheet to alpha-helix occurs essentially at the boundary between exons 2 and 3. Exons 1 and 2 therefore encode the putative globular domain(s), while exons 3 to 13 encode the putative helical tail of DIS1.

10 We propose that DIS1 should be considered as candidate gene involved in the aetiology of psychiatric disorders because it is directly disrupted by the translocation. In support of this contention is the predicted structure of DIS1, which is compatible with a role in development and functioning of the nervous system. The information contained herein, now enables the skilled person to assess the gene as candidate in

15 psychotic individuals unrelated to members of the family carrying the translocation. This is particularly important given that our mapping of the chromosome 1 breakpoint region has identified several ESTs which indicates the possible presence of additional genes. Even if such genes are not directly disrupted by the translocation, positional effects on their expression cannot be ruled out. Determination of the

20 genomic structure of DIS1 has provided the information required to look for mutations

in all of the transcribed sequence plus splice sites and DIS1 can now be evaluated by means of mutation screening and association studies.

The sequences of the present invention can be used to derive primers and probes for  
5 use in DNA amplification reactions in order to perform diagnostic procedures or to identify further, neighbouring genes which also may contribute to the development of schizophrenia.

It is known in the art that genes may vary within and among species with respect to  
10 their nucleotide sequence. The DIS1 genes from other species may be readily identified using the above probes and primers. Therefore, the invention also comprises functional equivalents, which are characterised in that they are capable of hybridising to at least part of the DIS1 sequence shown in SEQ ID NO: 1, preferably under high stringency conditions.

15 Two nucleic acid fragments are considered to have hybridisable sequences if they are capable to hybridising to one another under typical hybridisation and wash conditions, as described, for example in Maniatis, et al., pages 320-328, and 382-389, or using reduced stringency wash conditions that allow at most about 25-30%  
20 basepair mismatches, for example: 2x SSC, 0.1% SDS, room temperature twice, 30 minutes each, then 2x SSC, 0.1% SDS 37 °C once, 30 minutes; then 2X SSC, room temperature twice ten minutes each. Preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches. These degrees of homology can be selected by using wash conditions  
25 of appropriate stringency for identification of clones from gene libraries or other sources of genetic material, as is well known in the art.

Furthermore, to accommodate codon variability, the invention also includes sequences coding for the same amino acid sequences as the sequences disclosed  
30 herein. Also portions of the coding sequences coding for individual domains of the expressed protein are part of the invention as well as allelic and species variations thereof. Sometimes, a gene expresses different isoforms in a certain tissue which

includes splicing variants, that may result in an altered 5' or 3' mRNA or in the inclusion of an additional exon sequence. Alternatively, the messenger might have an exon less as compared to its counterpart as exemplified in the sequences enlisted here (SEQ ID NO: 3 contains an additional 22 amino acids as compared to SEQ ID 5 NO 2 due to an alternative splicing event). These sequences as well as the proteins encoded by these sequences all are expected to perform the same or similar functions and form also part of the invention.

The sequence information as provided herein should not be so narrowly construed 10 as to require inclusion of erroneously identified bases. The specific sequence disclosed herein can be readily used to isolate further genes which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors. Thus, in one aspect, the present invention provides for isolated polynucleotides 15 encoding a novel gene, disrupted in schizophrenia.

15

The DNA according to the invention may be obtained from cDNA. Alternatively, the coding sequence might be genomic DNA, or prepared using DNA synthesis techniques. The polynucleotide may also be in the form of RNA. The polynucleotide 20 may be in single stranded or double stranded form. The single strand might be the coding strand or the non-coding (anti-sense) strand.

The present invention further relates to polynucleotides which have at least 80%, 25 preferably 90% and more preferably 95% and even more preferably at least 98% identity with SEQ ID NO:1. Such polynucleotides encode polypeptides which retain the same biological function or activity as the natural, mature protein.

The percentage of identity between two sequences can be determined with programs such as DNAMAN (Lynnon Biosoft, version 3.2). Using this program two sequences can be aligned using the optimal alignment algorithm of Smith and Waterman (1981, 30 J. Mol. Biol., 147:195-197). After alignment of the two sequences the percentage identity can be calculated by dividing the number of identical nucleotides between the two sequences by the length of the aligned sequences minus the length of all gaps.

The DNA according to the invention will be very useful for *in vivo* or *in vitro* expression of the novel gene according to the invention in sufficient quantities and in substantially pure form.

5

In another aspect of the invention, there are provided polypeptides comprising the amino acid sequence encoded by the above described DNA molecules.

Preferably, the polypeptides according to the invention comprise at least part of the 10 amino acid sequences as shown in SEQ ID NO:2 and SEQ ID NO:3.

Also functional equivalents, that is polypeptides homologous to SEQ ID NO: 2 or SEQ ID NO: 3 or parts thereof having variations of the sequence while still maintaining functional characteristics, are included in the invention.

15

The variations that can occur in a sequence may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. Amino acid substitutions that are expected not to essentially alter biological and immunological 20 activities, have been described. Amino acid replacements between related amino acids or replacements which have occurred frequently in evolution are, *inter alia* Ser/Ala, Ser/Gly, Asp/Gly, Asp/Asn, Ile/Val (see Dayhof, M.D., *Atlas of protein sequence and structure*, Nat. Biomed. Res. Found., Washington D.C., 1978, vol. 5, suppl. 3). Based on this information Lipman and Pearson developed a method for 25 rapid and sensitive protein comparison (Science, 1985, 227, 1435-1441) and determining the functional similarity between homologous polypeptides. It will be clear that also polynucleotides coding for such variants are part of the invention.

The polypeptides according to the present invention include the polypeptides 30 comprising SEQ ID NO:2 and SEQ ID NO:3 but also their isoforms, i.e. polypeptides with a similarity of 70%, preferably 90%, more preferably 95%. Also portions of such polypeptides still capable of conferring biological effects are included. Especially

portions which still bind to ligands form part of the invention. Such portions may be functional per se, e.g. in solubilized form or they might be linked to other polypeptides, either by known biotechnological ways or by chemical synthesis, to obtain chimeric proteins. Such proteins might be useful as therapeutic agent in that

5 they may substitute the gene product in individuals with aberrant expression of the DIS1 gene.

The sequence of the gene may also be used in the preparation of vector molecules for the expression of the encoded protein in suitable host cells. A wide variety of host

10 cell and cloning vehicle combinations may be usefully employed in cloning the nucleic acid sequence coding for the DIS1 gene of the invention or parts thereof. For example, useful cloning vehicles may include chromosomal, non-chromosomal and synthetic DNA sequences such as various known bacterial plasmids and wider host range plasmids and vectors derived from combinations of plasmids and phage or

15 virus DNA.

Vehicles for use in expression of the genes or a ligand-binding domain thereof of the present invention will further comprise control sequences operably linked to the nucleic acid sequence coding for a ligand-binding domain. Such control sequences

20 generally comprise a promoter sequence and sequences which regulate and/or enhance expression levels. Of course control and other sequences can vary depending on the host cell selected.

Suitable expression vectors are for example bacterial or yeast plasmids, wide host

25 range plasmids and vectors derived from combinations of plasmid and phage or virus DNA. Vectors derived from chromosomal DNA are also included. Furthermore an origin of replication and/or a dominant selection marker can be present in the vector according to the invention. The vectors according to the invention are suitable for transforming a host cell.

Recombinant expression vectors comprising the DNA of the invention as well as cells transformed with said DNA or said expression vector also form part of the present invention.

5 Suitable host cells according to the invention are bacterial host cells, yeast and other fungi, plant or animal host such as Chinese Hamster Ovary cells or monkey cells. Thus, a host cell which comprises the DNA or expression vector according to the invention is also within the scope of the invention. The engineered host cells can be cultured in conventional nutrient media which can be modified e.g. for appropriate 10 selection, amplification or induction of transcription. The culture conditions such as temperature, pH, nutrients etc. are well known to those ordinary skilled in the art.

The techniques for the preparation of the DNA or the vector according to the invention as well as the transformation or transfection of a host cell with said DNA or 15 vector are standard and well known in the art, see for instance Sambrook et al., *Molecular Cloning: A laboratory Manual*. 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

The proteins according to the invention can be recovered and purified from 20 recombinant cell cultures by common biochemical purification methods including ammonium sulfate precipitation, extraction, chromatography such as hydrophobic interaction chromatography, cation or anion exchange chromatography or affinity chromatography and high performance liquid chromatography. If necessary, also protein refolding steps can be included.

25

DIS1 gene products according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with DNA according to the invention or an expression vector comprising DNA according to the invention, said cells 30 expressing the DIS1 gene products according to the invention.

Alternatively also the DIS1 gene products according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the DIS1 gene products.

- 5 Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or inhibition of a functional response is measured.
- 10 Thus, the present invention provides for a method for identifying ligands for DIS1 gene products, said method comprising the steps of:
  - a) introducing into a suitable host cell a polynucleotide according to the invention,
  - b) culturing cells under conditions to allow expression of the DNA sequence
  - 15 c) optionally isolating the expression product
  - d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said DNA from step a);
  - e) establishing whether a ligand has bound to the expressed protein.
  - 20 f) Optionally isolating and identifying the ligand

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

- 25 The present invention thus provides for a quick and economic method to screen for therapeutic agents for the prevention and/or treatment of diseases related to schizophrenia. The method is especially suited to be used for the high throughput screening of numerous potential compounds.
- 30 Compounds which activate or inhibit the function of DIS1 gene products may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present invention.

Also within the scope of the invention are antibodies, especially monoclonal antibodies raised against the polypeptide molecule according to the invention. Such antibodies can be used therapeutically to inhibit *DIS1* gene product function and 5 diagnostically to detect *DIS1* gene products.

The invention furthermore relates to the use of the *DIS1* gene products as part of a diagnostic assay for detecting psychiatric abnormalities or susceptibility to psychiatric disorders related to mutations in the nucleic acid sequences encoding the *DIS1* 10 gene. Such mutations may e.g. be detected by using PCR (Saiki et al., 1986, *Nature*, 324, 163-166). Also the relative levels of RNA can be determined using e.g. hybridization or quantitative PCR technology. The presence and the levels of the *DIS1* gene products themselves can be assayed by immunological technologies such as radioimmuno assays, Western blots and ELISA using specific antibodies 15 raised against the gene products. Such techniques for measuring RNA and protein levels are well known to the skilled artisan.

The determination of expression levels of the *DIS1* gene products in individual patients may lead to fine tuning of treatment protocols.

20 Also, transgenic animals may be prepared in which the expression of the *DIS1* gene is altered or abolished.

### Legends to the figures

25 Figure 1 Alignment of sequence immediately flanking the breakpoints from the normal chromosome 1, der (1), der (11) and normal chromosome 11 (wt1, der (1), der (11) and wt11 respectively).

30 Figure 2 Map of the chromosome 1 breakpoint region containing *DIS1*. Black boxes, *DIS1* exons; Letters marking vertical arrows, position of ESTs. Positions of the putative CpG island, putative translation start and stop sites, polyadenylation signals

and alternative splice site are indicated. EST accession numbers: A=AA777274, B=AA361879, C=AA311762, D=Hs.96883, E=AA249072, F=W04811, G=D78808, H=N49833, I=W29023/AA093172, J&K=H71071/Z40262, M=AA610789, 13=Hs.26985. ESTs J and K are located extremely close together such that their 5 order could not be determined.

## Examples

### 10 Example 1

#### Cloning of the Chromosome 1 Translocation Breakpoint

We have previously described the isolation of a 2.5kb *EcoRI* fragment (wt11) containing the normal chromosome 11 translocation breakpoint, and demonstrated that it hybridises to *EcoRI* fragments of 2.7kb and 7kb from the der (1) and der (11) 15 chromosomes respectively (1998, Psychiatr. Genet. 8: 175-181). This chromosome 11 breakpoint fragment was subcloned, and used to prepare a 2.15kb *HindIII/EcoRI* repeat-free sub-fragment with which an *EcoRI* total digest genomic library made from a cell line from a translocation carrier (MAFLI, 1993, Am. J. Hum. Genet. 52: 478-490) was screened. A 2.7kb *EcoRI* fragment, presumed to correspond to the der (1) 20 translocation fragment was obtained. This was confirmed by its hybridisation pattern (figure 1), where it hybridises to a 2.7kb fragment from MIS7.4, a hybrid cell line carrying the der (1) chromosome as its human component (1998, Psychiatr. Genet. 8: 175-181). Three fragments are visible from MAFLI; the 2.5kb wild-type 11 breakpoint fragment; the 2.7kb der (1) fragment and a fragment of 7.3kb, assumed to 25 be the wild-type chromosome 1 breakpoint fragment. This was confirmed using normal control human DNA which also shows hybridisation of the probe to the 2.5kb chromosome 11 breakpoint fragment, and to a 7.3kb fragment which must therefore be from chromosome 1.

The 2.7kb der (1) fragment was used to rescreen the library, avoiding any clones 30 which had previously hybridised to the chromosome 11 breakpoint fragment, and this yielded a 7.3kb clone (wt1), corresponding to the chromosome 1 breakpoint fragment.

## Example 2

### Identification of the Breakpoint

The cloned wt11, wt1 and der (1) fragments were sequenced and the positions of the 5 translocation breakpoints were identified by comparisons between these three sequences. Primers designed from wt11 and wt1 sequence amplified a 1.4kb fragment containing the breakpoint from the der (11) chromosome by PCR, and the product was partially sequenced. An alignment of the breakpoint sequences from each of the four chromosomes is presented in figure 1. This shows that the 10 translocation event resulted in no rearrangement at all on the der (1) chromosome, and a small rearrangement on the der (11), where there has been a deletion of the nucleotides TCAG accompanied by insertion of AA. This breakpoint sequence and minor rearrangement has been confirmed by genomic sequence analysis of two other translocation carriers (data not shown). The position of the breakpoint has also 15 been confirmed using pairs of primers, one primer pair from each side of the breakpoint, for PCR on genomic DNA from MIS7.4 and MIS39, cell lines carrying the der(1) and der (11) chromosomes respectively (data not shown).

## Example 3

### Breakpoint Sequence Analysis

The sequences of the breakpoint fragments from chromosomes 1 and 11 were used 20 to search sequence databases using BLAST (1997, Nucleic. Acids. Res. 25: 3389-3402) to identify matches indicating the presence of a gene, and also analysed using the suite of gene recognition and analysis programmes encompassed by Nucleotide 25 Identify X (NIX, <http://menu.hgmp.mrc.ac.uk/menu-bin/Nix/Nix.pl>).

BLAST searches of sequence databases identified sequence from one end of a BAC clone (Genbank/EMBL accession number AQ105798) within the wt11 fragment, but nothing else of note. Neither did NIX convincingly predict any exons to be present within the chromosome 1.1 breakpoint sequence. However the wt1 fragment contains 30 several interesting sequences. There is a tandemly repeated TAA trinucleotide which is contained within three overlapping sequence tagged sites (Genbank/EMBL accession numbers G09671, G09453 and G07779). These correspond to the marker

*D1S1621*, which maps approximately 120bp below the breakpoint. There are also sequence matches to the ends of three different BAC clones (Genbank/EMBL accession numbers AQ112950, AQ078498 and B40542).

From Genbank and EMBL, sequence matches to three separate expressed sequence tags (ESTs), and a messenger RNA, are also contained within the *wt1* fragment, all distal to the breakpoint. These are AA249072 (which overlaps with *D1S1621*), W04811, D78808 and AB007926, mapping approximately 80bp, 1.8kb, 2.8kb and 3.7kb from the breakpoint respectively (figure 2).

Homology to AA249072 and W04811 extends across the whole sequence obtained from each cDNA. However sequence corresponding to *wt1* in D78808 could be spurious. Only 103 nucleotides of the total 350 in the EST sequence are contained within the *wt1* sequence, yet this short match does not apparently correspond to an exon since there are no flanking splice sites. The remaining sequence is homologous to several other ESTs (UniGene cluster Hs.31446, <http://www.ncbi.nlm.nih.gov/UniGene/index.html>), none of which contain any *wt1* sequence or are even present on chromosome 1, as judged by a lack of hybridisation to genomic DNA from the chromosome 1 human/mouse hybrid cell line A9(Neo-1)-4 (data not shown). AB007926 consists of 6833 nucleotides of a brain-expressed transcript from chromosome 1 (1997, DNA Res. 4: 345-349). Only 189 nucleotides of this transcript are coincident with *wt1*.

NIX identified one putative exon with consensus splice sites on the forward strand of *wt1*. This exon contains all of the sequence match to mRNA AB007926. The match ends at the predicted splice sites, demonstrating the accuracy of the prediction.

## 25 Example 4

### Contig Construction

Genomic clones from the region were isolated from a PAC library, RPCI1 (1996, Construction of bacterial artificial chromosome libraries using the modified P1 (PAC) System. In "Current Protocols in Human Genetics", N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman and D. R. Smith, Eds., Unit 5.15 Pub. John Wiley and Sons, New York) distributed by the United Kingdom Human Genome Mapping Project Resource Centre, and a chromosome 1

cosmid library, provided by the Resource Centre of the German Human Genome Project at the Max-Planck-Institute for Molecular Genetics (1994, *Nature*, 367: 489-491, 1999, *Nature Genetics*, 22: 22). Contig construction essentially required three phases. Initially, genomic clones were identified by screening libraries with sequence

5 flanking the breakpoint, microdissection clone MD258 (1995, *Cytogenet. Cell Genet.* 70: 35-40), or with several cDNA fragments from DIS1. Overlaps between the clones were then determined by end sequencing using oligonucleotides bordering the cosmid and PAC vector cloning sites (data not shown). Pairs of primers were designed from the resulting sequence and overlapping clones were identified by PCR

10 (data not shown). For verification, the PCR products were hybridised to Southern blots of digested PAC and cosmid DNA (data not shown). Finally, remaining gaps in the contig were filled by further rounds of library screening using PCR products generated from clone ends. In addition, cosmid ICRFc112B0519Q6 was used to screen the PAC library to extend the contig in the proximal direction. Two markers,

15 *D1S251* and *D1S1621*, have been mapped on this contig. *D1S251* was mapped by PCR, while the location of *D1S1621* immediately distal to the breakpoint was determined by genomic sequencing. The locations of DIS1 exons 1-3 and 5-13 and of all the ESTs with respect to the cosmids and PACs were determined by hybridisation of oligonucleotides (not shown) to digested cosmid and PAC DNA.

20 ESTs 10 and 11 are located extremely close together such that their order with respect to the contig could not be determined by hybridisation. DIS1 exon 4 is known to be present in cosmid ICRFc112D2299QD4, but was not otherwise mapped because of the apparent presence of numerous related sequences in the surrounding DNA.

25

### **Example 5**

#### **A Contig Spanning the Chromosome 1 Translocation Breakpoint**

To investigate the genomic structure of DIS1 we first constructed a contiguous clone map spanning the chromosome 1 breakpoint (Fig. 1). This contig is estimated to

30 extend across at least 400kb based on average PAC and cosmid sizes of 130kb and 35kb respectively. Cosmid fluorescence *in situ* hybridisation to the translocation cell line MAFLI was employed to confirm the orientation of the contig, and that it crosses

the translocation breakpoint. Cosmids spanning the breakpoint, and located distal and proximal were found to hybridise as predicted. Cosmid ICRFc112I0142Q6 hybridises to the normal chromosome 1, and the derived 1 and derived 11 chromosomes, indicating that it crosses the breakpoint. Hybridisation of cosmid 5 ICRFc112D1274QD4 to the normal chromosome 1 and derived 1, shows that it is located proximal to the breakpoint. Finally, signal from cosmid ICRFc112G1395QD4 is visible on the normal chromosome 1 and the derived chromosome 11 demonstrating that this cosmid lies distal to the breakpoint.

10 **Example 6**

**Genomic Structure of DIS1**

Direct cosmid sequencing using primers designed from DIS1 cDNA sequence was used to determine the intron/exon structure of DIS1. Resulting genomic sequence was aligned with cDNA sequence and splice sites identified at the points of 15 divergence (table 1). Exons 1-3 and 5-13 were identified by this method. For technical reasons, exon 4 proved more difficult and splice site sequences were eventually identified by subcloning a genomic fragment containing the exon from a cosmid, followed by sequencing.

DIS1 consists of 13 exons extending across at least 300kb of genomic DNA (Fig. 1). 20 A region of 66 nucleotides which is deleted from some DIS1 transcripts was found to arise from utilisation of an internal splice donor site within exon 11 and the normal splice acceptor site of the same exon. The final intron of DIS1 is a member of the extremely rare AT-AC class of introns (1997, Trends. Biochem. Sci. 22:132-137). This intron has the consensus 5' and 3' splice site sequences, atatcctt and yccac 25 respectively, as well as the consensus branch-site element, tccttaac, close to the 3' splice site as shown in table 1. All the other introns are of the common class I type.

**Example 7**

**Mapping of Additional Transcribed Sequences in the Region**

30 During contig construction, all of the sequences generated from the ends of the PACs and cosmids, miscellaneous sequences and the sequence of ICRFc112I0142Q6, were used to screen Genbank and EMBL in search of

homologies to expressed sequence tags (ESTs). The locations of the 8 ESTs identified by database screening are shown (Fig. 1). Unigene cluster Hs.26985 (13) is derived from the 3' UTR of DIS1, while the remaining 8 ESTs have not yet been assigned to any known gene.

5

### Example 8

#### Expression of DIS1

When hybridised to Northern blots, DIS1 was found to be present in all adult human tissues examined, as a transcript of approximately 8.1kb. Various smaller transcripts

10 hybridise to the same probe. Although these may represent DIS1 splice variants, their significance is not yet known. In agreement with the Northern blot data, RT-PCR using primers towards the 5' end of DIS1 on a range of human foetal tissues also detected transcripts in every tissue tested (table 2).

Sample	age (weeks)	DIS1 proximal	DIS1 distal
brain	8.3	+	+ (2)
	10.3	+	+ (2)
	13.3	+	+ (2)
heart	8.0	+	+ (2)
	8.8	+	+ (2)
	9.1	+	+ (2)
	9.3	+	+ (2)
liver	10.6	+	+ (2)
kidney	10.0	+	+ (2)
spleen	14.8	+	+ (2)
limb	10.3	+	+ (2)

15 Table 2 RT-PCR analysis of DIS1 on a range of human foetal tissues. Approximate ages of gestation are given in weeks. 2: two bands obtained using one primer pair, +: transcript detected.

**Example 9****Tissue-specific distribution of DIS1**

Analysis of DIS1 expression indicates that the gene is widely expressed in foetal tissues, and that DIS1 transcripts are present in all adult tissues examined. However, 5 as well as normal functioning, it is also necessary to study what effect the translocation may have had on overall expression of the gene. DIS1 is disrupted within the open reading frame which may cause (1) production of a truncated transcript and protein retaining only one of the putative leucine zippers, (2) silencing of the disrupted allele, or (3) production of a fusion transcript/protein from a gene on 10 chromosome 11.

**Example 10****Cell Culture**

The lymphoblastoid cell line MAFL1 from an individual bearing the 15 t(1;11)(q42.1;q14.3) translocation, somatic cell hybrids MIS7.4 and MIS39 bearing the der (1) or der (11) translocation chromosomes respectively, and their culture conditions, have been described previously (1993, Am. J. Hum. Genet. 52: 478-490). Der (1) refers to the derived chromosome 1 where DNA from 1q42.1-qter has been lost and replaced with chromosome 11 material from 11q14.3-qter. Der (11) refers to 20 the reciprocal derived chromosome 11. The cell line A9(Neo-1)-4, a mouse A9 hybrid cell line carrying human chromosome 1, and its culture requirements, have been previously reported (1989, Jpn. J. Cancer Res., 80: 413-418).

**Example 11****25 PCR analysis of the breakpoint region of DIS1**

A 1.4kb product was amplified from the der 11 chromosome using one primer specific for chromosome 11 proximal to the breakpoint (ggctggatattgcccttgagccataatt) and one primer specific for chromosome 1 distal to the breakpoint (agaacagaggaggacgtatgac). This product was obtained using the cell line MIS39 30 which carries the der 11 chromosome. This product is only obtainable from the translocated chromosome.

**Example 12****FISH analysis of the breakpoint region of DIS1**

Cosmid fluorescence *in situ* hybridisation to the translocation cell line MAFLI was employed to confirm that the contig crosses the translocation breakpoint. Cosmid

5 ICRFc112I0142Q6 hybridises to the normal chromosome 1, and the derived 1 and derived 11 chromosomes, indicating that it crosses the breakpoint.

**Example 13****Methods****10 Fluorescence *in situ* Hybridisation**

Cosmids were mapped in relation to the chromosome 1 breakpoint using 2-7 day old slides of metaphase chromosomes prepared from the translocation cell line MAFLI by conventional methods. Cosmid DNA was labelled with dUTP-biotin by standard nick translation. FISH was carried out essentially as previously described (1995,

15 Genomics 28: 420-428). Slides were examined on a Leitz microscope and suitable metaphases scanned with a BioRad MRC-600 confocal laser scanning system.

**DNA Preparation**

Cosmid and PAC DNA was prepared by standard methods (Sambrook et al.,

20 *Molecular Cloning: A laboratory Manual*. 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).

Prior to sequencing, cosmid and PAC DNA was subjected to a phenol/chloroform clean-up step, followed by ethanol precipitation. Alternatively, cosmid DNA was prepared using Qiagen plasmid midi kits, followed by dialysis. Cosmid DNA prepared  
25 for sequencing was stored at 4°C. Plasmid DNA was prepared using QIAGEN plasmid midi kits.

**DNA Sequencing**

Cosmid end sequencing was carried out using primers 928

30 (aggcgcagaactggtaggtatg) and 929 (gctaaggatggttctagcgatg). PAC sequencing was carried out using primers SP6 (tactgtttgcgatctgccgtt) and T7 (aatacgactcactatagggaga). For cosmids and PACs 0.5-1 microgrammes of DNA was

sequenced using ABI PRISM Big Dye terminator cycle sequencing ready reaction kits with 60ng of primer. Plasmid DNA sequencing reactions were performed using ABI PRISM dRhodamine terminator cycle sequencing ready reaction kits and the products separated on an ABI 377 DNA sequencer (PE Applied Biosystems),  
5 according to the manufacturers instructions. Resulting sequence was analysed using the GCG package of sequence analysis software (Wisconsin package version 9.1, Genetics Computer Group, Madison, Wisc.). BLAST (1997, Nucleic. Acids. Res. 25: 3389-3402) searches were carried out at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

10

#### **RNA Extraction and cDNA Synthesis**

Human foetal tissues were obtained from the Medical Research Council Tissue Bank. Total RNA was extracted using RNazol B™ (Biogenesis Ltd.) according to the manufacturers instructions. First strand cDNA synthesis was carried out on DNase I  
15 treated RNA using the random hexamer primer from the SUPERSCRIPT™ Preamplification System (GIBCO BRL) according to the manufacturers instructions. 1 microlitre of the resulting cDNA was used in standard PCR reactions.

#### **Subcloning the Chromosome 11 Breakpoint Fragment**

20 The 2.5kb *EcoRI* fragment isolated as described previously (1998, Psychiatr. Genet. 8: 175-181) was cloned into *EcoRI*-digested pBluescript SK (-) (Stratagene) using standard methods (Sambrook et al., *Molecular Cloning: A laboratory Manual*, 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).

#### **25 Genomic Library Construction and Screening**

Genomic DNA from the translocation cell line MAFLI was digested with *EcoRI*, ligated into *EcoRI*-digested and dephosphorylated lambda ZAP II (Stratagene), and packaged using Gigapack Gold II packaging extract (Stratagene) according to the manufacturers instructions. Bacteriophage were plated using *E. coli* XL1-Blue MRF'  
30 and the library of clones screened using standard methods. Excision of clones from the lambda vector was carried out as advised by the manufacturer, releasing genomic fragments cloned into pBluescript SK (-). The library was screened using

the 2.15kb repeat-free *Hind*III/*Eco*RI fragment containing the chromosome 11 breakpoint, followed by the 2.7kb der (1) fragment. Of  $1 \times 10^7$  clones screened, one copy of the 2.5kb chromosome 11 fragment, four copies of the 2.7kb der (1) fragment, one copy of the 7.3kb chromosome 1 fragment and no copies of the 7kb 5 der (11) fragment were obtained.

### cDNA Library Screening

20-26 week foetal brain and 20-25 week foetal heart 5'-STRETCH PLUS cDNA libraries, constructed in lambda gt10 and gt11 respectively, were obtained from 10 Clontech and screened according to the manufacturers instructions. Inserts were obtained from pure clones using two methods. Firstly, cDNAs were amplified by PCR, turbocloned (1993, Nucleic Acids Res. 21: 817-821) and sequenced. Due to the probable introduction of sequence alterations during PCR, several subclones were sequenced. Alternatively, lambda DNA was digested with *Eco*RI to release the 15 cDNA insert which was then subcloned into *Eco*RI-digested pBluescript SK (-) (Stratagene).

### Polymerase Chain Reactions

PCR was carried out using AmpliTaq DNA polymerase (Perkin Elmer). Each 50 20 microlitre reaction contained 1 unit of enzyme, 300ng of each primer, 200mM of each dNTP, 1.5mM MgCl<sub>2</sub>, 50mM KCl and 10mM Tris-HCl pH8.3. A probe corresponding to nucleotides 1177-1321 of DIS1 was prepared from cloned cDNA using primers ACGTTACAACAAAGATTAGAAGACCTGG and TGCTGAGTGGCCCCACGGCGCAAG, with touchdown PCR (75°C-65°C) and 30s 25 denaturation at 94°C, 30s synthesis at 72°C. Marker D1S251 was mapped by PCR using the standard cycling conditions for this marker.

A probe containing the DIS1 exon predicted by NIX was prepared by PCR using the wt1 fragment as template and primers CCATTCTGGACGGCTAAAGACC & 30 GCARACACTTGGCTAAGGCAGGC (694bp product). The cycling conditions used were: 35 cycles of: 94°C, 30s; 58°C, 60s; 72°C, 60s. Amplification from DIS1 cDNA was performed using proximal primers CCAGAGCGTGACATGCATTC &

CCAGGTCTTCTAATCTTGTTGTAACGT (292bp product from 35 cycles of: 94°C, 30s; 62°C, 60s; 72°C, 30s) and distal primers GGAAGCTTGTGATTGCTTATCC & AGATCTTCATCATGACTGTGGATTGC (270 & 336bp products from 35 cycles of: 94°C, 30s; 64°C, 60s; 72°C, 30s). An initial hot start step was carried out. This 5 involved preparation of two separate mixes, one containing template, buffer and nucleotides, and the other containing enzyme and primers. These were incubated at 90°C separately for two minutes prior to mixing and cycling.

In order to amplify cDNA inserts from lambda vectors, a single plaque was picked into 25 microlitres of distilled water. 1-5 microlitres were then added to a PCR 10 reaction and the cDNA insert amplified using vector-based primers. Lambda gt10-specific primers, AGCAAGTTCAGCCTGGTTAAGT & GGGACCTTCTTATGAGTATT (35 cycles of: 94°C, 30s; 68°C, 60s; 72°C, 180s) and lambda gt11-specific primers GAAGGCACATGGCTGAATATCGACGGTTTC & GACACCAGACCAACTGGTAATGGTAGCGAC (35 cycles of 94°C, 30s; 56°C, 60s; 15 72°C, 90s) were used to amplify inserts from the foetal brain and foetal heart cDNA libraries respectively.

### Hybridisation

Standard procedures were used for Southern blotting and hybridisation (Sambrook et 20 al., *Molecular Cloning: A laboratory Manual*. 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). Probes were labelled with alpha <sup>32</sup>P-dCTP by random priming using High Prime (Boehringer Mannheim) and purified using Pharmacia NICK columns. The oligonucleotide probe was labelled with gamma <sup>32</sup>P-dATP. Oligonucleotide hybridisations were carried out overnight at the appropriate 25 temperature.

### Subcloning

Exon 4 of DIS1 (and flanking DNA) was subcloned from cosmid ICRFc112D2299QD4 by digestion with *EcoRI*. Digested fragments were subcloned 30 into *EcoRI*-digested pBluescript SK (-) (Stratagene) and subclones containing the exon were identified by hybridisation with the DIS1 cDNA nucleotide 1177-1321 probe. The exon was found to be contained within a fragment of approximately 4kb.

**Claims**

- 1 A substantially pure polynucleotide, encoding the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 3 or their isoforms.  
5
- 2 Polynucleotide according to claim 1, comprising the sequence according to SEQ ID NO: 1.
- 3 A recombinant expression vector comprising the polynucleotide according to claim 1 or 2 or fragments thereof.  
10
- 4 A polypeptide according to SEQ ID NO: 2 or SEQ ID NO: 3 or their isoforms.
- 5 Cell line transformed with a polynucleotide encoding at least part of the polypeptide according to claim 4  
15
- 6 Cell line transformed with a polynucleotide according to claim 1 or 2 or fragments thereof or transformed with the expression vector of claim 3.  
20
- 7 Cell line according to claim 6 of mammalian origin
- 8 Cell line according to claim 6 or 7 expressing a DIS1 gene product, wherein DIS1 gene is defined as a stretch of DNA
  - 25 a) of approximately 300 kB on chromosome 1, spanning the breakpoint of a balanced t(1;11)(q42.1;q14.3) translocation and
  - b) hybridisable to the polynucleotide sequence according to SEQ ID NO: 1 and/or SEQ ID NO: 4
- 9 Use of a polynucleotide hybridisable to the DIS1 gene in the in vitro diagnosis of  
30 a psychiatric disorder, wherein DIS1 gene is defined as a stretch of DNA
  - a) of approximately 300 kB on chromosome 1, spanning the breakpoint of a balanced t(1;11)(q42.1;q14.3) translocation and

b) hybridisable to the polynucleotide sequence according to SEQ ID NO: 1 and/or SEQ ID NO:4

10 10 Use of a cell line according to claim 6 to 8 in the in vitro diagnosis of a psychiatric disorder.

5

11 11 Use of a polypeptide encoded by a polynucleotide comprising SEQ ID NO 1 or fragments thereof in the in vitro diagnosis of a psychiatric disorder.

10 12 Use of a polynucleotide according to claims 1 or 2 or fragments thereof or the expression vector of claim 3 in a screening assay for the identification of new drugs.

15 13 Use of a polypeptide according to claim 4 or analogues or fragments thereof in a screening assay for the identification of drugs for the treatment of psychiatric disorders.

20 14 Use of a cell line according to claims 6 to 8 in a screening assay for the identification of new drugs for the treatment of psychiatric disorders.

15 15 A polynucleotide comprising SEQ ID NO 1 or fragments thereof for use as a medicament.

25 16 16 A polypeptide encoded by a polynucleotide comprising SEQ ID NO 1 or fragments thereof for use as a medicament

17 17 A polynucleotide comprising SEQ ID NO 1 or fragments thereof for use as a medicament for the treatment of a psychiatric disorder.

30 18 18 A polypeptide encoded by a polynucleotide comprising SEQ ID NO 1 or fragments thereof for use as a medicament for the treatment of a psychiatric disorder

19 Use of a polynucleotide comprising SEQ ID NO 1 or fragments thereof in the preparation of a medicament for the treatment of a psychiatric disorder

5 20 Use of a polypeptide encoded by a polynucleotide comprising SEQ ID NO 1 or fragments thereof in the preparation of a medicament for the treatment of a psychiatric disorder

21 Antibodies against the polypeptide according to claim 4

10

22 Pair of oligonucleotide primers for the amplification of a region containing a breakpoint involved in a balanced t(1;11)(q42.1;q14.3) translocation of chromosome 1 and chromosome 11, wherein a first primer is hybridisable to chromosome 1 or chromosome 11 and is located 5' of the breakpoint and a second primer which is also hybridisable to chromosome 1 or chromosome 11 is located 3' of the breakpoint.

15

23 Pair of oligonucleotide primers according to claim 22 wherein at least one of said primers is capable of hybridising to a sequence according to SEQ ID NO 4.

20

24 Pair of oligonucleotide primers according to claim 22 wherein at least one of said primers comprises the sequence of SEQ ID NO: 5 or SEQ ID NO: 6.

25 Method for the detection of a mutation in the DIS1 gene in a given subject comprising the steps of

- a) providing a set of oligonucleotide primers capable of hybridising to the nucleotide sequence of the DIS1 gene
- b) obtaining a sample containing nucleic acid from the subject
- c) amplifying a region flanked by the primer set of step 1 using a nucleic acid amplification method
- d) detecting whether the amplified region contains a mutation by

30

e) comparing the amplified sequence with the sequence of normal control subjects.

wherein DIS1 gene is defined as a stretch of DNA

- of approximately 300 kB on chromosome 1, spanning the breakpoint of a balanced t(1;11)(q42.1;q14.3) translocation and
- hybridisable to the polynucleotide sequence according to SEQ ID NO: 1 and/or SEQ ID NO:4

26 A method for identifying ligands for DIS1 gene products, said method comprising  
10 the steps of:

- a) introducing into a suitable host cell a polynucleotide according to claims 1 or 2 or an expression vector according to claim 3 or fragments thereof,
- b) culturing cells under conditions to allow expression of the DNA sequence
- c) optionally isolating the expression product
- 15 d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said DNA from step a);
- e) establishing whether a ligand has bound to the expressed protein.
- f) Optionally isolating and identifying the ligand

20

27 Method for the detection of a balanced t(1;11)(q42.1;q14.3) translocation in a given subject comprising the steps of

- a) providing a set of oligonucleotide primers according to claim 22
- b) obtaining a sample containing nucleic acid from the subject
- 25 c) amplifying a region flanked by the set of oligonucleotide primers of step a) using a nucleic acid amplification method
- d) detecting whether the translocation has occurred by
- e) comparing the amplified sequence with the sequence of normal control subjects.

30

28 Method according to claim 27 wherein the set of oligonucleotide primers comprises at least one primer comprising SEQ ID NO: 5 or SEQ ID NO: 6,

## BREAKPOINT

TGATGGGAGGCAACTTTCT~ der (11)

wt1

~ACACCGTCATCATCCTGTGAACCATGAACT

TGATGGGAGGCAACTTTCT~ wt1

der (1)

~ACACCGTCATCATCCTGTGAACCATGAACT

der(11)~ATTITTTAAATGATACTGAGATATCATGAA

wt11 ~ATTITTTAAATGATACTGAGATATCATGAACT

TCATTCTCCCTCATTATTCA~ wt11

TCATTCTCCCTCATTATTCA~ der(1)

Figures 1/2

Fig. 1

## Figures 2/2

Fig. 2

